Chemical Physics Letters 674 (2017) 173–178

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Frontiers article

Measuring molecular motions inside single cells with improved analysis of single-particle trajectories

David J. Rowland, Julie S. Biteen *

Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, United States

ARTICLE INFO

Article history:
Received 6 February 2017
In final form 15 February 2017
Available online 24 February 2017

Keywords:
Single-molecule microscopy
Super-resolution imaging
Cumulative probability distribution
Heterogeneous diffusion
Analysis algorithms
Parameter redundancy

ABSTRACT

Single-molecule super-resolution imaging and tracking can measure molecular motions inside living cells on the scale of the molecules themselves. Diffusion in biological systems commonly exhibits multiple modes of motion, which can be effectively quantified by fitting the cumulative probability distribution of the squared step sizes in a two-step fitting process. Here we combine this two-step fit into a single least-squares minimization; this new method vastly reduces the total number of fitting parameters and increases the precision with which diffusion may be measured. We demonstrate this Global Fit approach on a simulated two-component system as well as on a mixture of diffusing 80 nm and 200 nm gold spheres to show improvements in fitting robustness and localization precision compared to the traditional Local Fit algorithm.

1. Introduction

Subcellular dynamics vary in time and over small size scales due to spatial and temporal variations such as transient interactions with molecular partners [1], crowding by the nucleoid in bacteria [2], or the presence of different lipid domains in membranes [3]. To measure the motions of biological molecules such as proteins and lipids in the face of this complicated local environment, single-molecule super-resolution fluorescence microscopy data must be analyzed with a method that systematically accounts for heterogeneity. One approach to single-particle tracking (SPT) derives the apparent diffusion coefficient from each individual trajectory from the mean square displacement (MSD) vs. time lag curve [4–6]. The trajectory of each particle is thus assigned an average diffusion coefficient with this MSD approach, and heterogeneous diffusion is described by dividing the collection of MSD curves into diffusive populations. The number of trajectories in each population of curves may be taken as an estimate of the relative proportions of these diffusive populations [7–9]. However, in real systems, heterogeneous diffusion can be observed even over the course of the trajectory of a single molecule, and this single-track MSD analysis specifically disallows the case where a single molecular trajectory experiences multiple diffusive modes by providing only the average diffusion coefficient for each track.

An approach that accounts explicitly for such heterogeneous motion considers the entire collection of single-molecule steps instead of dividing these steps into individual tracks. This collection of step data can then be quantified based on the cumulative probability distribution (CPD) of the collection of squared step sizes to explicitly account for spatial and temporal heterogeneities, and increase the signal-to-noise ratio. Single-step CPD analysis is therefore a diffusion estimation technique that has had impact across fields by characterizing diverse biological systems such as artificial membranes, leukocytes, bacterial membranes, neurons and artificial materials [10–22].

Alternatively, a number of Bayesian [23–26] and machine learning algorithms [27] can be used to estimate the number of diffusive components and measure their properties, but the complexity of these methods poses a significant barrier to intuitive understanding of the underlying modes of heterogeneous motion. Fluorescence correlation spectroscopy (FCS) and the related methods of spatiotemporal image correlation spectroscopy (STICS), raster image correlation spectroscopy (RICS), or particle image correlation spectroscopy (PICS) can also quantify diffusion; these approaches all employ spatial or temporal correlation functions which can also be fit to multi-component diffusion models [28–32], but it is rare for the signal-to-noise to be high enough for the analysis of complex heterogeneous motion such as are found in bacterial systems [33]. The ability of single-particle tracking to isolate high quality trajectories from noisy single molecule data can present a more attractive conduit for analysis.

* Corresponding author.
E-mail address: jsbiteen@umich.edu (J.S. Biteen).

http://dx.doi.org/10.1016/j.cplett.2017.02.052
0009-2614/

Despite its advantages, the commonly employed CPD analysis method involves a two-step fitting process where the CPD and MSD curves are fit sequentially [11,13]. We present here an approach that combines this two-step Local Fit process into a single-step multi-domain Global Fit algorithm. We analyze simulated trajectories of multiple diffusive components to measure the improvements in the diffusion coefficient estimation error and find that Global Fit is superior to the traditional Local Fit CPD analysis algorithm. We then consider the diffusion of 80 nm and 200 nm gold spheres in glycerol water solution to show that Global Fit outperforms Local Fit in a real system. We report improvements in precision, robustness, and simplicity of use.

2. Materials and methods

2.1. Imaging and tracking

Slides were imaged at room temperature using wide-field epifluorescence microscopy in an Olympus IX71 inverted microscope with a 100×, 1.40 NA oil immersion objective (in Zeiss Immersol 518F immersion oil) and appropriate excitation, emission, and dichroic filters (Semrock LL01-488, Semrock BLP01-488 and Semrock Di01-R488, respectively). After a 3× beam expander, a Photometrics Evolve EMCCD camera with >90% quantum efficiency captured the images at 100 frames per second. Each camera pixel corresponds to a 49 nm × 49 nm area of the sample. The gold spheres were illuminated with a 488 nm laser (Coherent Sapphire 488-50), that was circularly polarized with a quarter waveplate (Tower Optical AO15Z 1/4). Single molecule positions were associated into tracks with the Hungarian algorithm [34] according to an exponential merit function [35].

2.2. Diffusion of gold spheres in glycerol

Gold nanoparticles with diameters 80 and 200 nm (BBI Solutions) were dispersed in 50% glycerol. 5 μL of the mixture was sandwiched between two glass coverslips. The second through fifth time lags were used for both the Global Fit and Local Fit algorithms to reduce the magnitude of the fitting residuals. Unweighted least squares fitting was performed with the Matlab built-in function *lsqnonlin*.

2.3. Simulations

Diffusion was simulated by generating 10^3 steps from a zero-mean normal distribution with variance equal to $2Dt_{\text{frame}}$, where D is the desired diffusion coefficient and t_{frame} is the simulated camera exposure time which was set to 0.04 s. Localization precision was simulated by adding zero-mean Gaussian-distributed random numbers to the simulated trajectories; the localization precision, or the standard deviation of the random numbers, was varied from 4.9 nm to 73.5 nm. Each simulation was repeated 10^3 times. The first 10 time lags were used for both the Global Fit and Local Fit algorithms and unweighted least squares was performed with the Matlab built-in function *lsqnonlin*.

2.4. Bootstrapping

For the analysis of the tracks of gold spheres, histograms of estimated diffusion coefficients and population weights were produced by bootstrapping the fitting procedure. The total set of 13232 squared step sizes was sampled with replacement 300 times to produce 300 unique data sets each with as many values as the original data set. These bootstrapped data sets were then fit with either the Global Fit or Local Fit method.

2.5. CPD Local Fit

To probe heterogeneous diffusion, the cumulative probability distribution (CPD) of squared step sizes (Δr^2) was calculated from the tracks of diffusing molecules at each time lag (τ) between frames in the trajectory. There is one CPD curve, CPD_i, for each time lag considered, and each CPD_i was fit to the multi-term exponential fit [11] with the appropriate number of terms (three terms shown here for instance):

$$CPD_i = 1 - \alpha_1 \times \exp\left(\frac{-\Delta r_i^2}{MSD_{1,i}}\right) - \alpha_2 \times \exp\left(\frac{-\Delta r_i^2}{MSD_{2,i}}\right) - (1 - \alpha_1 - \alpha_2) \times \exp\left(\frac{-\Delta r_i^2}{MSD_{3,i}}\right) \quad (1)$$

This series of fits, where i, runs from 1 to the number of time lags considered, N_τ, estimates three mean squared displacements, MSD_1, MSD_2, and MSD_3, as a function of time lag for $N_D = 3$ diffusive populations with weights α_1, α_2, and $(1 - \alpha_1 - \alpha_2)$, respectively. Each of the three MSD curves is then fit to a model—here of 2D unconfined diffusion—to extract the diffusion coefficient of the respective population of molecules. For example, for population 1, the second-step fitting function is:

$$MSD_1 = 4D_1\tau + 4\sigma_1^2, \quad (2)$$

where D_1 is the diffusion coefficient of population α_1, τ is the domain of time lags and σ_1 is the localization precision for population α_1. If one uses the first 5 time lags to estimate the diffusion coefficients of three populations, the total number of fitting parameters in this local CPD fitting approach is: $5 \times N_\tau + 2 \times N_D = 31$ here with $N_\tau = 5$, and $N_D = 3$.

2.6. CPD Global Fit

Instead of fitting in separate steps, the set of empirical CPDs may be fit all at once by incorporating the MSD functions (Eq. (2)) into Eq. (1). Conceptually, this can be understood as the sharing of redundant parameters, such as the weight of population 1, i.e., α_1. The free parameters in the combined fitting function now include only N_D diffusion coefficients (one for each population), a single localization precision, σ, shared among all populations, and all but one of the population weights because one is estimated using the others. For instance, if one wishes to estimate the diffusion coefficients of three populations, the total number of fitting parameters is 6 (three diffusion coefficients, one localization precision parameter, and two population weight parameters). This number of fitting parameters in Global Fit is hugely improved from the 31 parameters necessary for Local Fit.

We implemented Global Fit with a Matlab-specific formulation that exchanges several nonlinear least squares problems for a single larger nonlinear least squares problem. See https://github.com/BiteenMatlab/SingleMoleculeDataAnalysis for complete code; the key snippets of our code describing the use of this algorithm for a three-population Global Fit are given here:

```
msdFun = @(tau,p)cat(2, 4*p(1)*tau + p(2),
  4*p(3)*tau + p(2), 4*p(4)*tau + p(2));


cpdFun = @(x,y,p)1-p(5)*exp(-x/y(1))-p(6)*exp(-x/y
(2))-(1-p(5)-p(6))*exp(-x/y(3));
  -p(6)*exp(-x/y(2))-(1-p(5)-p(6))
  *exp(-x/y(3));
```